BS a novel LINE-like element in Drosophila melanogaster

1354–1358 Nucleic Acids Research, 1995, Vol. 23, No. 8

BS a novel LINE-like element in *Drosophila melanogaster*

A. Udomkit, S. Forbes, G. Dalgleish and D. J. Finnegan*

Institute of Cell and Molecular Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JR, UK

Received December 30, 1994; Revised and Accepted March 7, 1995

EMBL accession no. X77571

ABSTRACT

Transposable elements with long terminal inverted repeats are rare and only one family of elements of this sort has been identified in the genome of *Drosophila melanogaster*. An insertion associated with the Hs^{BS} mutation of the *achaete–scute* complex has been reported to be a second element of this type. We have determined the complete sequence of this insertion and have shown that it is in fact two copies of a new LINE-like transposable element, that we have called *BS*, inserted in opposite orientation 337 bp apart. Like other elements of this type, *BS* has two open reading frames that appear to encode a *gag*-like polypeptide and a reverse transcriptase. There are few complete *BS* elements in the five strains of *D.melanogaster* that we have tested and they appear to transpose infrequently. The events that may have lead to the double *BS* insertion are discussed in terms of the supposed mechanism of transposition of LINE-like elements.

INTRODUCTION

Transposable elements are an important component of eukaryotic genomes, making up 15% or more of the total DNA. They occur as families of dispersed repeat sequences scattered throughout the genome, the number of copies varying from less than ten to several hundred thousand depending on the element and species concerned. They can be classified according to their structure and presumed mechanism of transposition, and fall into two main classes: elements that transpose by reverse transcription of an RNA intermediate, Class I elements, and elements that transpose directly from DNA to DNA, Class II elements.

There are two types of Class I elements. Class I.1 elements resemble retroviruses in having long terminal direct repeats (LTRs) and open reading frames with the potential to code for polypeptides similar to viral *gag*, *pol* and, in some cases, *env* proteins. Class I.2 elements also have open reading frames with similarities to retroviral *gag* and *pol* genes, but have no terminal repeats and end with A-rich sequences at the 3′ end of their coding strands. Elements of this type are often referred to as LINE-like elements as the first examples to be detected were mammalian LINE or L1 elements (1).

The majority of Class II elements have short terminal inverted repeats, but some have been reported to have inverted repeats that are hundreds of base pairs long. Few elements with this structure have been identified, the best characterized being the FB elements of *Drosophila melanogaster* (2). These have inverted repeats 1–2 kb long that are themselves made up of short tandem repeats. Some FB elements contain nothing more than these tandem repeats, while others have a central region of unrelated and non-repeated sequence (3–5). In one case this central region appears to be due to insertion of a different Class II element (6), while the central region of another may code for a function that stimulates FB transposition (7,8).

Campuzano *et al.* (9) reported finding what appeared to be a member of a second family of elements with long inverted repeats that they named *BS* because it was found associated with the mutation Hw^{BS}. This is a derivative of Hw^{1}, a mutation of the *achaete–scute* complex (AS–C) that is associated with an insertion of a copy of the retrovirus-like element *gypsy* within the T5 AS-C transcription unit. Chromosomes carrying the Hw^{BS} mutation have a slightly reduced hairy wing phenotype and an additional 8 kb of DNA inserted within the Hw^{1} *gypsy* element. This extra DNA contains inverted repeats of 2.5 kb as judged by restriction site mapping and hybridization data. The 8 kb insertion does not contain *FB* sequences and was named *BS*.

In order to compare the structure of this putative new inverted repeat element with that of *FB* elements, we have determined the nucleotide sequence of the whole of the *BS* insertion. The results indicate that the Hw^{BS} mutation is not associated with insertion of a single inverted repeat element, but has two related sequences inserted 337 bp apart and in opposite orientation within the *gypsy* element of Hw^{1}. These are members of a new family of LINE-like elements as they show no strong sequence similarity to any of the five families of such elements, *Doc*, *F*, *G*, *I* and *jockey*, that have been identified in the genome of *D.melanogaster* (10,11).

MATERIALS AND METHODS

DNA sequencing

The sequence of the Hw^{BS} insertion was determined by subcloning the *Pst*I fragments BS0.6, BS816, BS817, BS818 and BS819, and the *Pst*I–*Bam*HI fragment BS2.0 (Fig. 1A) of the plasmid p14RRHw^{BS} (9) in the vector pUBS (12). Double-stranded templates were prepared using QIAGEN™ columns and these were sequenced by dideoxynucleotide chain termination (13). The *Hin*dIII fragments BS1.1, BS5.8 and BS7.6 were used to orient and link the sequences from the *Pst*I fragments. The entire

* To whom correspondence should be addressed

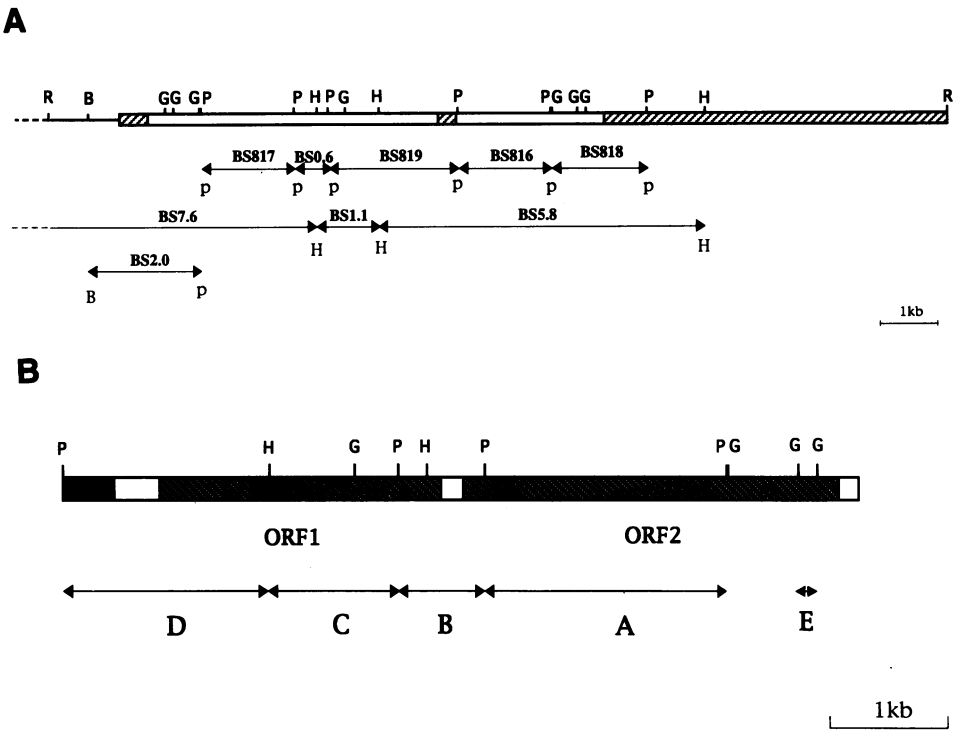

Figure 1. (**A**) Restriction map of an *Eco*RI fragment from the region of the AS-C region containing the Hw^{BS} insertions. The open boxes represent *BS* sequences. *BS*1 is on the left and *BS*2 on the right. The hatched boxes represent *gypsy* sequences, the thin lines represents AS-C sequences and the dashed line represents vector sequences. The lines underneath the map indicated fragments that were subcloned for DNA sequencing. (**B**) Restriction map of the *BS*1 insertion shown in the 5′→3′ orientation. The two open reading frames are indicated by hatched boxes while non-coding regions are indicated by open boxes. The filled box represents *gypsy* DNA. The lines underneath the map indicate restriction fragments used as probes for the Southern hybridizations shown in Figure 4. Restriction sites are indicated as follows: B, *Bam*HI; G, *Bgl*I; H. *Hin*dIII; P, *Pst*I; R, *Eco*RI.

sequence has been determined on both strands of the element. Sequence data were assembled using programs written by Staden (14) and were analysed using the suite of programs from the Wisconsin Genetics Computer Group (15). The program PILEUP is from Version 7.0 of the GCG package (Genetics Computer Group, Inc.).

Drosophila DNA preparation and Southern blotting

Genomic DNA was prepared from strains of *Drosophila* by homogenizing 10 flies in homogenization buffer (8 M urea, 0.35 M NaCl, 10 mM Tris–HCl pH 8.3, 10 mM EDTA). Homogenates were extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1) and then with chloroform:isoamyl alcohol (24:1). Nucleic acids were ethanol precipitated and dissolved in 50 ml of 20 mg/ml RNase in 10 mM Tris–HCl pH 8.3, 1 mM EDTA.

Five micrograms of DNA from each strain were digested with the appropriate restriction enzymes and electrophoresed on 0.7% agarose gels together with amounts of control plasmids equivalent to two, five, eight and 10 copies per genome. The DNA was then transferred to GeneScreen Plus™ nylon membrane (Dupont) by capillary action in 1.5 M NaCl, 0.15 M Na citrate. Hybridization was carried out in 0.5 M Na phosphate pH 7.2, 7% sodium dodecyl sulphate and 1 mM EDTA. After hybridization filters were washed in 40 mM Na phosphate, 1% sodium dodecyl sulphate, 1 mM EDTA. Fragments used as probe were purified by agarose gel electrophoresis and labelled with [^{32}P]adCTP by random priming (16). The amount of hybridization in each track was quantified with the aid of a PhosphorImager (Molecular Dynamics).

RESULTS

We have determined the sequence of the *BS* insertion starting from a 14.8 kb *Eco*RI fragment of Hw^{BS} genomic DNA kindly supplied by Dr S. Campuzano. This contains the entire *gypsy* and *BS* insertions associated with the Hw^{1} mutation (Fig. 1A). The sequence of the *BS* insertion is diagrammed in Figure 1B and is in the EMBL data base as accession X77571. It is not an insertion of a single inverted repeat element, as had been thought, but is made up of two copies of an element lying close together and in opposite orientation within *gypsy*. We have called the left-hand element *BS*1 and the right-hand element *BS*2.

The *BS*1 and *BS*2 insertions are 5126 and 2571 bp long, respectively, and each is flanked by a duplication of a 12 bp sequence present only once in the corresponding part of the *gypsy* sequence. These sequences are different for *BS*1 and *BS*2, and were presumably generated when these elements inserted at these sites. These insertions are separated by 337 bp of *gypsy* DNA. Geyer *et al.* (17) have determined what they believed to be the ends of the whole *BS* insertion when analysing insertions that modify the phenotype of mutations due to *gypsy* insertions. The sequences that they reported were those of the ends of *BS*2.

The structure of the *BS*1 element indicates that it is a LINE-like (Class I.2) transposable element. It has no terminal repeats, has an A-rich sequence at the 3′ end of the coding strand and has two long open reading frames (Fig. 1B). The first has three copies of

A

```
BS        CRRCQEYGHTAKYC
jockey    CQRCQIFGHSKNYC
Doc       CTNCQEYGHTKAYC
F         CTNCQEYGHTRGYC
G         CFRCQGFGHTQRYC
I         CKKCLRFGHPTPIC
R1Dm      CHRCVGFDHKVSEC
          *  *   *   *

BS        CARCGEN-HQTMQ--C
jockey    CGKCSGP-HMTGFALC
Doc       CVVCSEP-HTTAN--C
F         CVVCGDL-HDSKQ--C
G         CKVCGGL-HDSRA--C
I         CINCSETKHTNDGEKC
R1Dm      CRQCGQQ-HTAAK--C
          *  *   *      *

BS        CYHCSEN-HTASFKGC
jockey    CINCGGD-HVSTDKSC
Doc       CSNCGEK-HTANYRGC
F         CNNCGGN-HTANYRGC
G         CLHCQAD-HPASFKGC
I         CLNCRNNPELDHQHSP
R1Dm      CRNCRHRGQPSGHYML
          *  *   *      *
```

B

```
BS        YRPISLLSSLSKLWER
jockey    YRPTSLLPSLGKIMER
Doc       YRPISLLTCLSKLFEK
F         YRPISLLSCISKLFEK
G         YRPISLLAILSKILER
I         YRPISLNCCIAKILDK
R1Dm      YRPISLNCCIAKILDK
R2Dm      FRPISVPSVLVRQLNA
          *******  *  *  **

BS        LATYADDIALLYS
jockey    IATYADDTAVLTK
Doc       TSTFADDTAILSR
F         VSTFADDTAILSR
G         VATYADDTAFLAS
I         FNAYADDFFLIIN
R1Dm      LSAYADDLLLLVE
R2Dm      AAAFADDLVLFAE
            *****  *  *
```

Figure 2. (**A**) Alignment of the sequences of the first open reading frames of seven LINE-like elements from *D.melanogaster* showing the conserved CCHC motifs. Amino acid residues are represented by the standard one-letter code. Positions at which the same residue occurs in at least four of the sequences are marked with asterisks. (**B**) Alignments of amino acid sequences from parts of the second open reading frame of eight LINE-like elements of *D.melanogaster*. The first alignment shows the conserved motif YRPISLL that is characteristic of LINE-like elements (35,36) and YXDD that is found in all reverse transcriptases (37). Sequences corresponding to the other conserved motifs of reverse transcriptases are also present in ORF2 of *BS*. Positions at which the same residue is found in at least five of the sequences are marked with an asterisk. The sequences were obtained from the following *jockey* (38), *Doc* (39), *F* (40), *G* (41), *I* (42), *R1Dm* and *R2Dm* (20).

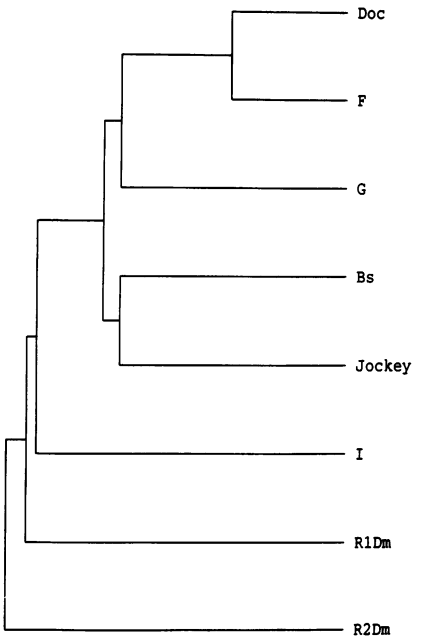

Figure 3. Dendrogram showing the relationships between the amino acid sequences of the putative reverse transcriptases encoded by the second open reading frames of the LINE-like transposable elements of *D.melanogaster* shown in Figure 2B. Distances along the horizontal axis are proportional to the differences between sequences.

a CCHC motif like those found in the *gag* genes of retroviruses and in the first open reading frames of LINE-like elements (Fig. 2A). The second has motifs characteristic of a reverse transcriptase (Fig. 2B). Five other transposable elements of this type, *Doc*, *F*, *G*, *I* and *jockey*, have been found in the genome of *D.melanogaster* (10,11), but the sequence of *BS*1 has not been reported previously. Elements of this type are frequently truncated by varying amounts at their 5′ ends (1,18). This is thought to happen at some point during transposition (19). The sequence of *BS*2 is identical to *BS*1, but is missing the first 2555 bp, suggesting that it has suffered such a deletion.

We have determined the relationships between the amino acid sequences of the putative reverse transcriptases of *BS*1 and the five *Drosophila* LINE-like elements mentioned above as well as those of two other potential transposable elements of this type, R1Dm and R2Dm (20), that are found inserted at specific sites within a proportion of the 28S rRNA coding sequences of *D.melanogaster*. These are believed to differ from the other LINE-like elements in being able to insert at these sites in a sequence-specific manner. This was done by comparing the complete amino acid sequence encoded by the second open reading frame of each element using the GCG program PILEUP. The result (Fig. 3) suggests that *BS* is most closely related to *jockey*.

We have used Southern transfer experiments to measure the number of *BS* elements in the genomes of five different stocks of *D.melanogaster* and to estimate their degree of conservation (Fig. 4). Three restriction fragments of *BS*1 were used to probe digests of genomic DNA using enzymes that should excise the corresponding fragments from complete genomic copies. Digests of a plasmid containing the fragment used as probe were used as size markers and to give an indication of the number of copies per genome. The intensity of hybridization of fragments corresponding in size to the probe decreases from the 3′ to 5′ end of the element (Fig. 4A–C) confirming that *BS* elements are a family of sequences containing variable 5′ deletions.

There are about five copies of the 3′ most fragment tested (fragment A in Fig. 1B) in the strains tested. The higher molecular weight bands presumably correspond to *BS* elements truncated somewhere within fragment A. This gives an upper limit of about 10 to the number of *BS* elements in the genome. The results from using fragment C as probe suggest that the strains tested contain few if any complete BS elements.

Campuzano *et al.* (9) estimated the copy number of *BS* elements to be 15 using the 1.7 kb *Pst*I fragment (fragment D in Fig. 1B) as probe. This fragment includes both internal *gypsy* sequences and sequences from the 5′ end of *BS* so this estimate reflects the number of copies of both elements. We were unable to detect fragments corresponding to this 1.7 kb *Hin*dIII–*Pst*I fragment in DNA of any of the five strains tested (Fig. 4D), indicating that the double insertion of *BS* elements within *gypsy* does not occur in any of these strains.

We have compared the distribution of *BS* elements in the genomes of five strains of *D.melanogaster* by probing *Pst*I-digested DNA with a *Bgl*II fragment from the 3′ end of *BS*1. Each *BS* element should give a band of hybridization that depends on its position in the genome. The patterns of hybridization are similar from strain to strain (Fig. 5) suggesting that *BS* elements transpose infrequently in them. Unfortunately we have not been able to obtain the Hw^{BS} strain to check that it does not have an unusually high frequency of *BS* transposition.

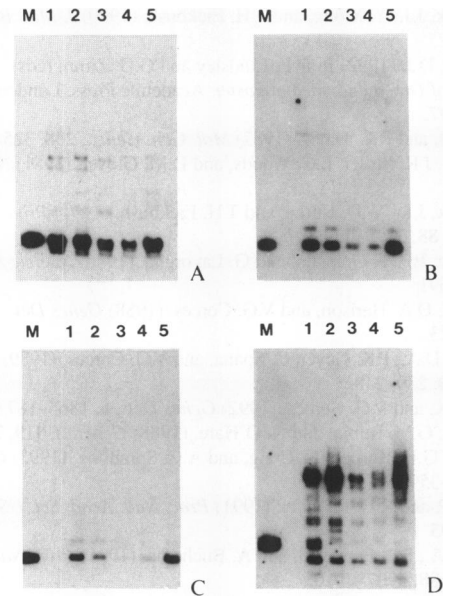

Figure 4. Southern transfer experiment showing the conservation of *BS* elements within the genomes of five strains of *D.melanogaster* isolated from the wild. Genomic DNAs were digested with *Pst*I in (**A**) and (**B**) and with *Pst*I and *Hin*dIII in (**C**) and (**D**). The fragments used to probe these DNAs are shown in Figure 1B. In each case a digest of a plasmid including the fragment used as probe was loaded on the gel in varying amounts to serve as a size marker and to allow the amount of hybridization of the genomic DNAs to be quantified. Track M shows the plasmid DNA loaded in an amount equivalent to 15 copies per haploid genome. The strains of *D.melanogaster* are as follows; 1, w^k; 2, Samarkand; 3, Charolles; 4, Oregon R; 5, Canton S.

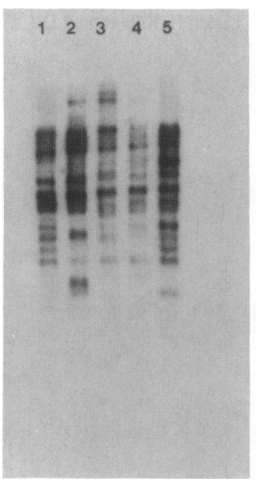

Figure 5. Southern transfer experiment showing the degree of heterogeneity in the distribution of *BS* elements in the genomes of the strains of *D.melanogaster* analysed in Figure 4. Genomic DNAs were digested with *Pst*I and probed with the *Bgl*II fragment of *BS*, Fragment E in Figure 1B.

DISCUSSION

The insertion associated with the Hw^{BS} mutation is not an inverted repeat element like *FB*, as had been supposed previously, but is formed by two copies of a new element, *BS*, inserted in opposite orientation. These elements have all the characteristics of LINE-like elements and *BS*1 may be full-length.

Although the proportion of the *D.melanogaster* genome that comprises transposable elements is not unusually large, the number of different transposable elements of all kinds that can be found in this species is greater than that known for any other eukaryote (21). Eight LINE-like elements have now been identified in *D.melanogaster* as well as a short sequence that may represent a ninth family (22). This is the greatest number of elements of this type to have been found in any species. Five of these, *Doc F, I, jockey* and, presumably, *BS*, can insert at many sites in the genome while two of them, *R1Dm* and *R2Dm*, are site-specific elements (23,24). *G* elements are concentrated in the non-transcribed spacer DNA of ribosomal gene clusters and may not be active in most strains (25).

The *BS*2 element has inserted within one copy of a tandemly repeated 12 bp sequence that is the binding site for the product of the *suppresser of Hairy wing* gene (26). Binding of su(Hw) protein to these repeats in a *gypsy* element inserted upstream of the *yellow* gene alters *yellow* expression by interfering with the interaction of the *yellow* promoter and some of its upstream regulatory sequences (27,28). Insertion of the transposable elements *hobo* or *jockey* within these repeats suppresses the mutant phenotype of y^2 presumably because this reduces the binding of su(Hw) protein (17).

The *gypsy* element associated with Hw^1 is inserted within the transcribed region of the T5 (achaete) gene of the AS–C and the mutant allele is transcribed to give a hybrid T5–*gypsy* transcript that is about 10-fold more abundant than T5 transcripts from the wild-type allele (9). The level of this hybrid transcript is reduced in flies carrying the *su(Hw)* mutation suggesting that in this case binding of su(Hw) protein may prevent interaction of the T5 promoter and a *cis*-acting negative regulator. The binding of su(Hw) protein to its target may be reduced somewhat by insertion of the *BS*2 element within one of the 12 bp repeats as the Hw^1 phenotype is slightly attenuated in Hw^{BS} flies, although no effect on the level of the T5-*gypsy* transcript has been detected (9).

The Hw^{BS} insertion must have been produced by either two independent transposition events or a single transposition event that generated two copies of *BS*. Insertions of *P* elements have been found that are close together and in opposite orientation (29). These are probably the results of independent insertions as *P* elements transpose at high frequency and preferentially to sites adjacent to a donor element (30). Elements that transpose by reverse transcription of an RNA intermediate, as is the case for LINE-like elements (19,31,32), are unlikely to transpose preferentially to adjacent sites and the frequency of *BS* transposition appears to be low. This suggests that the Hw^{BS} double insertion is the result of a single transposition event.

LINE-like elements are thought to integrate by a mechanism in which the 3′-OH at a nick in a strand of chromosomal DNA is used to prime DNA synthesis by an element encoded reverse transcriptase that uses the RNA transposition intermediate as template (33,34). The DNA synthesized in this way is in turn used as the template for synthesis of the second strand to complete integration. If the RNA intermediate were to be released after synthesis of the first DNA strand it might to serve as primer for a second integration event close by resulting in a double insertion.

ACKNOWLEDGEMENTS

This work was supported by research grants from the Medical Research Council and the Human Frontier Science Program and a studentship to AU from the Government of Thailand. We are grateful to S. Campuzano for supplying the original *BS* clone.

REFERENCES

1 Hutchison III, C.A., S.C. Hardies, D.D. Loeb, W.R. Sehee, and M.H. Edgell. (1989) In D.E. Berg and M.M. Howe (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 593–617.
2 Bingham, P.M. and Z. Zachar. (1989) In D.E. Berg and M.M. Howe (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 485–502.
3 Truett, M.A., R.S. Jones, and S.S. Potter. (1981) *Cell*, **24**, 753–763.
4 Potter, S.S. (1982) *Mol. Gen. Genet.*, **188**, 107–110.
5 Potter, S.S. (1982) *Nature*, **297**, 201–204.
6 Brierley, H.L. and S.S. Potter. (1985) *Nucleic Acids Res.*, **13**, 485–500.
7 Templeton, N.S. and S.S. Potter. (1989) *EMBO J.*, **8**, 1887–1894.
8 Harden, N. and M. Ashburner. (1990) *Genetics*, **126**, 387–400.
9 Campuzano, S., L. Balcells, R. Villares, L. Carramolino, L. Garcoa-Alonso, and J. Modolell. (1986) *Cell*, **44**, 303–312.
10 Finnegan, D.J. (1989) In D.E. Berg and M.M. Howe (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 519–522.
11 Finnegan, D.J. (1989) In D.E. Berg and M.M. Howe (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 503–517.
12 Goman, M., B. Mons, and J. Scaife. (1991) *Mol. Biochem. Parasitol.*, **45**, 281–288.
13 Sanger, F., S. Nicklen, and A.R. Coulson. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
14 Staden, R. (1982) *Nucleic Acids Res.*, **10**, 4731–4751.
15 Devereux, J., P. Haeberli, and O. Smithies. (1984) *Nucleic Acids Res.*, **12**, 387–395.
16 Feinberg, A.P. and B. Vogelstein. (1983) *Anal. Biochem.*, **132**, 6–13.
17 Geyer, P.K., M.M. Green, and V.G. Corces. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8593–8597.
18 Bucheton, A. (1990) *Trends Genet.*, **6**, 16–21.
19 Jensen, S. and T. Heidmann. (1991) *EMBO J.*, **10**, 1927–1937.
20 Jakubczak, J.J., Y. Xiong, and T.H. Eickbush. (1990) *J. Mol. Biol.*, **212**, 37–52.
21 Finnegan, D.J. (1992) In D.L. Lindsley and G.G. Zimm (eds), *The Genome of Drosophila melanogaster*. Academic Press, London, pp. 1096–1107.
22 Pittler, S.J. and R.L. Davis. (1987) *Mol. Gen. Genet.*, **208**, 325–328.
23 Roiha, H., J.R. Miller, L.C. Woods, and D.M. Glover. (1981) *Nature*, **290**, 749–753.
24 Jakubczak, J.L., W.D. Burke, and T.H. Eickbush. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 3295–3299.
25 Di Nocera, P.P., F. Graziani, and G. Lavorgna. (1986) *Nucleic Acids Res.*, **14**, 675–691.
26 Spana, C., D.A. Harison, and V.G. Corces. (1988) *Genes Dev.*, **2**, 1414–1423.
27 Harrison, D.A., P.K. Geyer, C. Spana, and V.G. Corces. (1989) *Dev. Genet.*, **10**, 239–248.
28 Geyer, P.K. and V.G. Corces. (1992) *Genes Dev.*, **6**, 1865–1873.
29 Roiha, H., G.M. Rubin, and K. O'Hare. (1988) *Genetics*, **119**, 75–83.
30 Tower, J., G.H. Karpen, N. Craig, and A.C. Spradling. (1993) *Genetics*, **133**, 347–359.
31 Evans, J.P. and R.D. Palmiter. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 8792–8795.
32 Pélisson, A., D.J. Finnegan, and A. Bucheton. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 4907–4910.
33 Pritchard, M.A., J.-M. Dura, A. Pelisson, A. Bucheton, and D.J. Finnegan. (1988) *Mol. Gen. Genet.*, **214**, 533–540.
34 Schwartz-Sommer, Z., L. Leclercq, F. Gobel, and H. Saedler. (1987) *EMBO J.*, **6**, 3873–3880.
35 Xiong, Y. and T.E. Eickbush. (1990) *EMBO J.*, **9**, 3353–3362.
36 Finnegan, D.J. (1988) In *Transposition*. Cambridge University Press.
37 Toh, H., R. Kikuno, H. Hayashida, T. Miyata, W. Kugimiya, S. Inouye, S. Yuki, and K. Saigo. (1985) *EMBO J.*, **4**, 1267–1272.
38 Primagii, A.F., L.J. Mizrokhi, and Y.V. Ilyin. (1988) *Gene*, **70**, 253–262.
39 O'Hare, K., M.R.K. Alley, T.E. Cullingford, A. Driver, and M.J. Sanderson. (1991) *Mol. Gen. Genet.*, **225**, 17–24.
40 Di Nocera, P.P. and G. Casari. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5483–5487.
41 Di Nocera, P. (1988) *Nucleic Acids Res.*, **16**, 4041–4052.
42 Fawcett, D.H., C.K. Lister, E. Kellett, and D.J. Finnegan. (1986) *Cell*, **47**, 1007–1015.